Published in final edited form as:

Underst Complex Syst. 2011 ; 72: 485–494. doi:10.1007/978-3-642-20853-9_33.

Multiple comparison of trends in cancer rates taking into account overlapping cases*

Nirian Martín¹ and Yi Li²

¹ Dept. Statistics, Carlos III University of Madrid, Spain nirian.martin@uc3m.es

² Dept. Biostatistics – Harvard School of Public Health and Dep. Biostatistics & Computational Biology – Dana Farber Cancer Institute, USA yili@jimmy.harvard.edu

Summary

When analyzing trends in cancer rates, it is common to rely on the so-called Annual Percent Change (APC). For dealing with such a measure of trend, directly age-adjusted rates are usually considered. Classical methods such as pooled t-tests are often applied for comparing APCs of two groups of individuals in a simple way under independence assumption. In practice, it is quite common to find groups of interest for which the independence assumption fails because their regions or periods of time overlap. No one of the papers that deal this problem consider the case where more than two APCs are compared. In this work we propose a Wald-type test-statistic which is not difficult to compute once we provide the estimators of two or more APCs to be compared. These estimators are the minimum power divergence estimators that cover as special case those obtained by maximum likelihood.

1 Introduction

According to the World Health Statistics 2009, published by WHO, the different types of cancer jointly cardiovascular diseases constituted the main causes of death during year 2004, specially in upper-middle and high income countries. It is well known that as income level increases, the percentage of non-communicable deaths (NCDs) increases, in other words, as people live longer, the risk of NCD grows. In such a study it is shown that in high-income countries during year 2004, 77% of all the deaths were NCDs. Cardiovascular diseases and cancers, as main part of NCDs, provided values of age-standardized mortality rates equal to 408 per 100,000 and 164 per 100,000 respectively. In order to implement prevention and control plans against these diseases, it is very important for health authorities to analyze whether the mortality (incidence) rates related to such diseases have increased or decreased in a set of successive years and in what degree, that is the trends in mortality (incidence) are taken into account. The National Surveillance, Epidemiology and End Results (SEER [2]) Program of the National Cancer Institute (NCI) constitutes nowadays one of the most prestigious epidemiological surveillance program in the world that allows the scientific community to do research on cancer data based on the US population, actually the SEER provides a database and statistical software, SEER*Stat. The new techniques that are needed to cover different problems related to the age-adjusted cancer rates are being continuously implemented by the SEER and particularly the Annual Percent Change (APC) for studying trends of cancer rates is considered.

*This work is related to the stay of Nirian Martín in Harvard University (September 2008 - August 2009), supported by the Real Colegio Complutense. She met Marisa just the first day after arriving in Spain, September 1. In the last conversation Marisa told Nirian that both will meet soon again. Due to her optimistic character it was impossible to imagine they both could not meet again. This paper has been written in memory to Marisa.

Statistically, the trend in cancer rates is an average rate of change per year in a given period of time framework when constant change along the time has been assumed. Let r_i be the expected mean of the cancer rate associated with the i -th time point in a sequence of ordered I time points $\{t_i\}_{i=1}^I$, for example each time point could be a year ($t_i = i$, $i = 1, \dots, I$). The cancer rates can be referred to either cancer incidence rates or cancer mortality rates which demonstrate the risk of developing cancer or dying from cancer respectively. The annual percent change (APC) is a measure that allows us to compare trends in age-adjusted rates that are taken year by year. Regarding the age-adjusted rates, if we apply directly standardized rates, the expected mean of the cancer rate associated with the rate of the k -th region ($k = 1, \dots, K$) at the time-point t_{ki} ($i = 1, \dots, I_k$), or the i -th year ($t_{ki} = i$), is given by $r_{ki} = \sum_{j=1}^J \omega_j r_{kji}$, where J is the number of age-groups, $\{\omega_j\}_{j=1}^J$ is the age-distribution of the Standard Population ($\sum_{j=1}^J \omega_j = 1$, $\omega_j > 0$, $j = 1, \dots, J$) and $\{r_{kji}\}_{j=1}^J$ is the set of rates associated with the k -th region at the time-point i in each of the age-groups ($j = 1, \dots, J$). The SEER Program applies directly standardized rates on the US population of year 2000 with $J = 19$ age-groups [0, 1), [1, 5), [5, 10), [10, 15), ..., [80, 85), [85, *). The APC associated with the rates $\{r_{kji}\}_{j=1}^J$ of the k -th region is defined as the percentage

$$APC_k = 100 (\theta_k - 1), \quad (1)$$

where $\theta_k = r_{k2}/r_{k1} = \dots = r_{kI_k}/r_{k,I_k-1}$ (constant change assumption along the whole period of time in region k).

Let n_{kji} the population at risk in the k -th region, j -th age-group, at the time-point t_{ki} and d_{kji} the number of deaths (or incidences) in the k -th region, j -th age-group, at the time-point t_{ki} . The r.v.s that generate d_{kji} , D_{kji} , are considered to be mutually independent with Poisson distribution $D_{kji} \sim \mathcal{P}(m_{kji})$, where $m_{kji} \equiv E[D_{kji}] = n_{kji} r_{kji}$.

The Age-stratified Poisson Regression model, introduced by Li *et al.* [6], is useful for modeling $\{r_{kji}\}_{j=1}^J$ under constant change assumption. For m_{kji} (or for r_{kji}) it is considered to hold

$$\log \frac{m_{kji}}{n_{kji}} = \beta_{0kj} + \beta_{1kj} t_{ki} \quad \text{or} \quad \log r_{kji} = \beta_{0kj} + \beta_{1kj} t_{ki}. \quad (2)$$

The parameter of the model is $\beta_k \equiv (\beta_{0k1}, \dots, \beta_{0kJ}, \beta_{1k})^T$ and its parameter space $\Theta_k \equiv \mathbb{R}^{J+1}$. Note that

$$\theta_k = \exp(\beta_{1k}), \quad (3)$$

with β_{1k} being the parameter we would like to estimate in the practice. The same model in matrix notation is given by

$$\log (\text{Diag}^{-1}(\mathbf{n}_k) \mathbf{m}_k(\beta_k)) = \mathbf{X}_k \beta_k \quad \text{or} \quad \mathbf{m}_k(\beta_k) = \text{Diag}(\mathbf{n}_k) \exp(\mathbf{X}_k \beta_k), \quad (4)$$

where $\mathbf{m}_k(\beta_k) \equiv (m_{k11}(\beta_k), \dots, m_{kI_k}(\beta_k))^T$, $\text{Diag}(\mathbf{n}_k)$ is a diagonal matrix of individuals at risk $\mathbf{n}_k \equiv (n_{k11}, \dots, n_{kI_k})^T$ and

$$\mathbf{X}_k = \begin{pmatrix} 1_{I_k} & & \mathbf{t}_k \\ & \ddots & \vdots \\ & & 1_{I_k} & \mathbf{t}_k \end{pmatrix}_{J I_k \times (J+1)} = (\mathbf{I}_J \otimes 1_{I_k}, 1_J \otimes \mathbf{t}_k),$$

with $\mathbf{t}_k \equiv (t_{k1}, \dots, t_{kJ})^T$, is a full rank $M_k \times (J+1)$ design matrix, with $M_k \equiv J I_k$.

If we focuss on model (4), for making statistical inference with sample $\mathbf{D}_k \equiv (D_{k11}, \dots, D_{kJ I_k})^T$ there is no problem. However, if we take jointly $k = 1, \dots, K$ regions, $\mathbf{D} = (\mathbf{D}_1, \dots, \mathbf{D}_K)^T$ is not a simple random sample, because it is possible to find individual at risk at the same time in different regions. In order to overcome this difficulty in Martín and Li [3] it were made some assumptions with regard to the overlapping region associated exclusively to a couple of regions k and h $k \neq h$, which is applied in the case where both regions have the same APC: if $D_{kji}^{(h)} \sim \mathcal{P}(m_{kji}^{(h)})$ is the number of deaths (incidences) in the overlapping part “in space” between regions k and h , then for $n_{kji}^{(h)} > 0$ (population at risk) it holds

$$m_{kji}^{(h)} = \frac{n_{kji}^{(h)}}{n_{kji}} m_{kji}. \quad (5)$$

In addition, in the complementary part with respect to the overlapping part “in space” between regions k and h , it holds $D_{kji}'^{(h)} \sim \mathcal{P}(m_{kji}'^{(h)})$, which is independent with respect to $D_{kji}^{(h)}$ and $m_{kji}'^{(h)} = (n_{kji}'^{(h)} / n_{kji}) m_{kji}$.

In this paper the most important characteristics for the minimum power divergence estimators of the Age-stratified Poisson Regression model, when overlapping group of individuals are taking into account, are revised in Section 2. As main part of the paper Wald-type test-statistics for testing the equality of APCs of K regions with possible overlapping group of individuals are presented in Section 3. A numerical result is shown in Section 4.

2 Minimum power divergence estimators for the age-stratified Poisson regression model with a possible overlapping group of individuals

Based on the likelihood function of a Poisson sample associated exclusively to the k -th region, $\mathbf{D}_k$, and using a single index for vectors rather than triple index, the kernel of the log-

likelihood function is given by $\ell_{\beta_k}(\mathbf{D}_k) = \sum_{s=1}^{M_k} D_s \log m_s(\beta_k) - \sum_{s=1}^{M_k} m_s(\beta_k)$, and thus the MLE of β_k is $\hat{\beta}_k = \arg \max_{\beta_k \in \Theta_k} \ell_{\beta_k}(\mathbf{D}_k)$. Focussed on a multinomial contingency table it is intuitively understandable that a good estimator of the probabilities of the cells should be such that the discrepancy with respect to the empirical distribution or relative frequencies is small enough. The oldest discrepancy or distance measure we know is the Kullback divergence measure, actually the estimator which is built from the Kullback divergence measure is the MLE. By considering the unknown parameters of a Poisson contingency table, the expected values, rather than probabilities and the observed frequencies rather than relative frequencies, we are going to show how is it possible to carry out statistical inference for Poisson models through power divergence measures. According to the Kullback divergence measure, the discrepancy or distance between the Poisson sample $\mathbf{D}_k$ and its vector of means $\mathbf{m}_k(\beta_k)$ is given by

$$d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = \sum_{s=1}^{M_k} \left(D_s \log \frac{D_s}{m_s(\beta_k)} - D_s + m_s(\beta_k) \right). \quad (6)$$

Observe that $d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = -\ell_{\beta_k}(\mathbf{D}_k) + C(k)$, where $C(k)$ does not depend on parameter β_k . Based on such a relationship we can define the MLE of β_k as the minimum Kullback divergence estimator $\widehat{\beta}_k = \arg \min_{\beta_k \in \Theta_k} d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k))$, and the MLE of $\mathbf{m}_k(\beta_k)$ functionally as $\mathbf{m}_k(\widehat{\beta}_k)$, due to the invariance property of the MLEs. The power divergence measures are a family of measures defined as

$$d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = \frac{1}{\lambda(1+\lambda)} \sum_{s=1}^{M_k} \left(\frac{D_s^{\lambda+1}}{m_s^\lambda(\beta_k)} - D_s(1+\lambda) + \lambda m_s(\beta_k) \right), \quad (7)$$

such that from each possible value for subscript $\lambda \in \mathbb{R} - \{0, -1\}$ a different way to quantify the discrepancy between $\mathbf{D}_k$ and $\mathbf{m}_k$ arises. In case of $\lambda \in \{0, -1\}$, it is defined $d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = \lim_{\ell \rightarrow \lambda} d_\ell(\mathbf{D}_k, \mathbf{m}_k(\beta_k))$, and in this manner the Kullback divergence appears as special case of power divergence measures when $\lambda = 0$, $d_0(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = d_{\text{Kull}}(\mathbf{D}_k, \mathbf{m}_k(\beta_k))$ and on the other hand case $\lambda = -1$ is obtained by changing the order of the arguments for the Kullback divergence measure, $d_{-1}(\mathbf{D}_k, \mathbf{m}_k(\beta_k)) = d_{\text{Kull}}(\mathbf{m}_k(\beta_k), \mathbf{D}_k)$. The estimator of β_k obtained on the basis of (7) is the so-called minimum power divergence estimator (MPDE) and it is defined for each value of $\lambda \in \mathbb{R}$ as

$$\widehat{\beta}_{k,\lambda} = \arg \min_{\beta_k \in \Theta_k} d_\lambda(\mathbf{D}_k, \mathbf{m}_k(\beta_k)), \quad (8)$$

and the MPDE of $\mathbf{m}_k(\beta_k)$ functionally as $\mathbf{m}_k(\widehat{\beta}_{k,\lambda})$ due to the invariance property of the MPDEs. For more details about such a estimator see Pardo and Martín [5].

In order to obtain the MPDE of (1), $\widehat{\text{APC}}_{k,\lambda} = 100 \left(\exp(\widehat{\beta}_{1k,\lambda}) - 1 \right)$, we need to compute the estimator of the parameter of interest by following the next result. In Martín and Li [3] it was established the next result.

Proposition 1. *The MPDE of β_{1k} , $\widehat{\beta}_{1k,\lambda}$, is the solution of the nonlinear equation*

$$f(\widehat{\beta}_{1k,\lambda}) = \sum_{i=1}^{I_k} t_{ki} Y_{ki} = 0,$$

with

$$\begin{aligned} Y_{ki} &= \sum_{j=1}^J m_{kji}(\widehat{\beta}_\lambda) (\varphi_{kji} - 1), \\ m_{kji}(\widehat{\beta}_\lambda) &= n_{kji} \exp(\widehat{\beta}_{0kj,\lambda}) \exp(\widehat{\beta}_{1ki,\lambda} t_{ki}) \quad \text{and} \quad \varphi_{kji} = \left(\frac{D_{kji}}{m_{kji}(\widehat{\beta}_\lambda)} \right)^{\lambda+1}, \\ \exp(\widehat{\beta}_{0kj,\lambda}) &= \left(\sum_{s=1}^{I_k} p_{kjs} \psi_{kjs}^{\lambda+1} \right)^{\frac{1}{\lambda+1}}, \quad j=1, \dots, J, \\ p_{kjs} &= \frac{n_{kjs} \exp(\widehat{\beta}_{1k,\lambda} t_{ks})}{\sum_{h=1}^{I_k} n_{kjh} \exp(\widehat{\beta}_{1k,\lambda} t_{kh})} \quad \text{and} \quad \psi_{kjs} = \frac{D_{kjs}}{n_{kjs} \exp(\widehat{\beta}_{1k,\lambda} t_{ks})}. \end{aligned}$$

For the asymptotic distributions some assumptions have to be made:

- a. $m_{kji}^*(\beta_k^0) = m_{kji}(\beta_k^0)/N_k$ remains constant as $N_k \equiv \sum_{s=1}^{M_k} m_s(\beta_k)$ increases, that is $m_{kji}(\beta_k^0)$ increases at the same rate as N_k .
- b. $N_k^* \equiv \frac{N_k}{N} (k=1, \dots, K)$ is constant as $N \equiv N_1 + \dots + N_K$ increases, that is N increases at the same rate as N_k .

For details about the asymptotic distribution of parameters see Martín and Li [3].

3 Wald-type test-statistics for the age-stratified Poisson regression model with a possible overlapping group of individuals

The Z -test provided in Li *et al.* [6],

$$Z_{\lambda}(r;s) = \frac{\widehat{\beta}_{1r,\lambda} - \widehat{\beta}_{1s,\lambda}}{\sqrt{\widehat{\text{Var}}(\widehat{\beta}_{1r,\lambda} - \widehat{\beta}_{1s,\lambda})}} \quad (9)$$

with $\lambda = 0$, for testing $\mathcal{H}_0(r;s): \text{APC}_r = \text{APC}_s$ (or $\mathcal{H}_0(r;s): \beta_{1r} - \beta_{1s} = 0$) based on the age-adjusted Poisson Regression model is useful only if we want to compare the APCs of two specific regions r and s . This test-statistic is not valid only for MLEs ($\lambda = 0$) but also for MPDEs, actually all the MPDEs of β have the same asymptotic distribution. The asymptotic distribution of $Z(r; s)$ is standard normal under $\mathcal{H}_0(r;s)$. Based directly on (9), in Martín and Li [3] $K - 1$ test hypotheses separately were performed, $\mathcal{H}_0(1;k): \text{APC}_1 = \text{APC}_k, k=2, \dots, K$. For the moment there is no paper concerned in making comparisons between more than two regions at the same time, $\mathcal{H}_0: \text{APC}_1 = \text{APC}_2 = \dots = \text{APC}_K$. In order to do that we reformulate the hypotheses

$$\mathcal{H}_0(1;2, \dots, K): \theta = (\theta_1, \dots, \theta_{K-1})^T = 0, \quad (10)$$

against $\mathcal{H}_1(1;2, \dots, K): \exists b \in \{1, \dots, K-1\}: \theta_b \neq 0$, where $\theta_b \equiv \beta_{11} - \beta_{1b+1}, b = 1, \dots, K-1$. The following test-statistic is applied

$$W_{\lambda}(1;2, \dots, K) = \widehat{\theta}_{\lambda}^T \widehat{\text{Var}}^{-1}(\widehat{\theta}_{\lambda}) \widehat{\theta}_{\lambda}, \quad (11)$$

$\widehat{\theta}_{\lambda} = (\widehat{\theta}_{1,\lambda}, \dots, \widehat{\theta}_{K-1,\lambda})^T$, where $\widehat{\theta}_{b,\lambda} \equiv \widehat{\beta}_{11,\lambda} - \widehat{\beta}_{1b+1,\lambda}, b=1, \dots, K-1$. In particular, there are no individuals at risk shared by regions to be compared, its expression is given by

$$W_{\lambda}(1;2, \dots, K) = \sum_{b=1}^{K-1} \frac{\widehat{\theta}_{b,\lambda}^2}{\widehat{\sigma}_{1,b+1,\lambda}^2} - \frac{1}{\sum_{k=1}^K \frac{1}{\widehat{\sigma}_{1k,\lambda}^2}} \sum_{b=1}^{K-1} \sum_{b'=1}^{K-1} \frac{\widehat{\theta}_{b,\lambda} \widehat{\theta}_{b',\lambda}}{\widehat{\sigma}_{1,b+1,\lambda}^2 \widehat{\sigma}_{1,b'+1,\lambda}^2}. \quad (12)$$

Note that in practice when $K = 2$ test-statistic (12) is equivalent to (9) with

$$\widehat{\text{Var}}(\widehat{\beta}_{1r,\lambda} - \widehat{\beta}_{1s,\lambda}) = \widehat{\sigma}_{1r,\lambda}^2 + \widehat{\sigma}_{1s,\lambda}^2 \text{ and } r = 1, s = 2, \text{ because } W_{\lambda}(1;2) = Z_{\lambda}^2(1;2).$$

Our aim in this paper is to find the asymptotic distribution of $\widehat{\theta}_{\lambda}$ first and the distribution of (11) later. When comparing regions b and b' , we must consider a new reference point for time index i , denoted by bb' , such that $t_{1bb'}$ represents the time point within $\{t_{1i}\}_{i=1}^{I_b}$ where

the time series associated with region b' is about to start, i.e. we have $\{t_{b'i}\}_{i=1}^{I_b}$ such that $t_{b1} = t_{1bb'} + 1$. Taking into account Martín and Li [3] it is easy to prove the next result.

Theorem 1. *Under the hypothesis that $\theta_b = 0$ or $\beta_{11} = \beta_{1b+1}$, $b = 1, \dots, K-1$, the asymptotic distribution of $\widehat{\theta}_\lambda$ is central Normal with*

$$\begin{aligned} \text{Var}(\widehat{\theta}_{b,\lambda}) &= \sigma_{11}^2 + \sigma_{1b+1}^2 - 2\sigma_{11}^2 \sigma_{1b+1}^2 \xi_{1,b+1}, \\ \text{Cov}(\widehat{\theta}_{b,\lambda}, \widehat{\theta}_{b',\lambda}) &= \sigma_{11}^2 + \sigma_{1b+1}^2 \sigma_{1b'+1}^2 \xi_{b+1,b'+1} - \sigma_{11}^2 \sigma_{1b+1}^2 \xi_{1b+1} - \sigma_{11}^2 \sigma_{1b'+1}^2 \xi_{1b'+1}, \end{aligned}$$

where σ_{1k}^2 is equal to

$$\sigma_{1k}^2 = \left(\sum_{j=1}^J \sum_{i=1}^{I_k} m_{kji}(\beta^0) (t_{ki} - \widetilde{t}_{kj}(\beta_k^0))^2 \right)^{-1}, \quad (13)$$

with $\widetilde{t}_{kj}(\beta_k^0) = \left(\sum_{i=1}^{I_k} m_{kji}(\beta_k^0) t_{ki} \right) / \sum_{i=1}^{I_k} m_{kji}(\beta_k^0)$ and

$$\xi_{bb'} = \sum_{j=1}^J \sum_{i=1}^{I_b - I_{bb'}} \frac{n_{b'ji}^{(b)}}{n_{b'ji}} m_{b'ji}(\beta^0) (t_{b'i} - \widetilde{t}_{bj}(\beta^0)) (t_{b'i} - \widetilde{t}_{b'j}(\beta^0)). \quad (14)$$

Proof. It follows by the bilinear property of covariance and taking into account the asymptotic distribution of $\widehat{\beta}_\lambda$ that was established in Martín and Li [3]: The asymptotic distribution of $\widehat{\beta}_\lambda$ is central Normal with $\text{Var}(\widehat{\beta}_{1k,\lambda}) = \sigma_{1k}^2$ and $\text{Cov}(\widehat{\beta}_{1b,\lambda}, \widehat{\beta}_{1b',\lambda}) = \sigma_{1b}^2 \sigma_{1b'}^2 \xi_{bb'}$.

We need to obtain the MPDEs of σ_{1k}^2 , $\xi_{bb'}$, $\widehat{\sigma}_{1k,\lambda}^2$ and $\widehat{\xi}_{bb',\lambda}$ respectively. A way to proceed is based on replacing β^0 by the most efficient MPDE. For example, for $\widehat{\sigma}_{1k,\lambda}^2$ we shall use

$$\widehat{\beta}_\lambda^0 \equiv \begin{cases} \widehat{\beta}_{1,\lambda}^0, & \text{if } N_1 \geq N_k \\ \widehat{\beta}_{k,\lambda}^0, & \text{if } N_1 < N_k \end{cases}.$$

Theorem 2. *The asymptotic distribution of $W_\lambda(1; 2, \dots, K)$ is χ_{K-1}^2 under (10).*

Proof. Let $\mathbf{L}^T = (\mathbf{1}_{K-1}, -\mathbf{I}_{K-1})$ with $\mathbf{1}_{K-1}$ being a $(K-1)$ -dimensional vector of 1's and $\mathbf{I}_{K-1}$ is the $(K-1)$ -dimensional identity matrix. Taking into account that $\widehat{\beta}_\lambda = \mathbf{L}^T \widehat{\theta}_\lambda$ and based on (14) and Corollary 2.1 in Dik and de Gunst [1] it is concluded that $W_\lambda(1; 2, \dots, K)$ is χ_g^2 under (10), where $g = \text{rank}(\mathbf{L}) = K-1$.

To determine whether N is large, it can be used the rule of thumb of the average of events per cell $\eta \equiv N/(KJ)$. Data sets with $\eta < 5$ are considered to be small, but the length of the whole period of time T is also an important factor to take into account. Because for $\lambda = 0$, $\widehat{N}_{k,0} = \sum_{j=1}^J \sum_{i=1}^{I_k} D_{kji}$, it is also held $\widehat{N}_0 = \sum_{k=1}^K \widehat{N}_{k,0} = \sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^{I_k} D_{kji}$. Hence in practice $\widehat{\eta} \equiv \sum_{k=1}^K \sum_{j=1}^J \sum_{i=1}^{I_k} D_{kji} / (KJ)$ is useful to estimate whether N is large.

4 Real data example

We have considered thyroid cancer mortality in three regions, Western (W) US population (compounded by Arizona, New Mexico and Texas), South Western (SW) US population (compounded by Arizona, California and Nevada) and West Coast (WC) US population (compounded by California, Oregon and Washington). Note that Arizona is shared by W and SW, and California by SW and WC. Different periods of time, 1969-1983, 1977-1991 and 1990-1999 are taken for W, SW and WC respectively. The third one differs from the rest in the sense that it considers a shorter period of time for its study. In Martín and Li [3] it was analyzed the performance of the minimum chi-square estimator (MCSE), that is the minimum power divergence estimator obtained with $\lambda = 1$, and it was concluded that it behaves more accurately than the MLEs ($\lambda = 0$) for small data sets. We have chosen thyroid cancer because it is a rare cancer, and both estimators have been computed in order to analyze the data. The rates are expressed per 100, 000 individuals at risk. Taking into

account $\exp(\beta_{0k}) = \sum_{j=1}^J \omega_j \exp(\beta_{0kj})$, in Figure 1 the fitted models are plotted and from them it seems at first sight that there is a decreasing trend for Thyroid cancer in WC and SW, and null or decreasing trend in W.

The specific values for estimates and test-statistics $W_\lambda(1; 2, 3)$, for $\lambda = 0, 1$, are summarized in Table 1.

Apart from the Wald-type test-statistic, we have included test-statistics $Z_{12,\lambda}$, $Z_{23,\lambda}$ ($\lambda = 0, 1$) for couples of regions. For their computation are essential the values of three matrices:

$$\begin{aligned} (\widehat{\xi}_{bb',0})_{b,b' \in \{1,2,3\}} &= \begin{pmatrix} * & -8684.09 & 0 \\ -8684.09 & * & 827.56 \\ 0 & -827.56 & * \end{pmatrix}, (\text{Cov}(\widehat{\beta}_{1b,0}, \widehat{\beta}_{1b',0}))_{b,b' \in \{1,2,3\}} \\ &= \begin{pmatrix} 2.92354 \times 10^{-5} & -0.77292 \times 10^{-5} & 0 \\ -0.77292 \times 10^{-5} & 3.04440 \times 10^{-5} & -0.32915 \times 10^{-5} \\ 0 & -0.32915 \times 10^{-5} & 13.06460 \times 10^{-5} \end{pmatrix}, \\ \text{Var}(\widehat{\theta}_0) &= \begin{pmatrix} 7.51380 \times 10^{-5} & 3.68340 \times 10^{-5} \\ 3.68340 \times 10^{-5} & 15.98815 \times 10^{-5} \end{pmatrix}. \\ (\widehat{\xi}_{bb',1})_{b,b' \in \{1,2,3\}} &= \begin{pmatrix} * & -9125.47 & 0 \\ -9125.47 & * & -866.18 \\ 0 & -866.18 & * \end{pmatrix}, (\text{Cov}(\widehat{\beta}_{1b,1}, \widehat{\beta}_{1b',1}))_{b,b' \in \{1,2,3\}} \\ &= \begin{pmatrix} 2.78598 \times 10^{-5} & -0.73429 \times 10^{-5} & 0 \\ -0.73429 \times 10^{-5} & 2.88823 \times 10^{-5} & -0.31074 \times 10^{-5} \\ 0 & -0.31074 \times 10^{-5} & 12.42113 \times 10^{-5} \end{pmatrix}, \\ \text{Var}(\widehat{\theta}_1) &= \begin{pmatrix} 7.14279 \times 10^{-5} & 3.50927 \times 10^{-5} \\ 3.50927 \times 10^{-5} & 15.20711 \times 10^{-5} \end{pmatrix}. \end{aligned}$$

Because $\text{p-value}(Z_{12,0}) = 0.064$, $\text{p-value}(Z_{12,1}) = 0.054$, $\text{p-value}(Z_{23,0}) = 0.39$, $\text{p-value}(Z_{23,1}) = 0.45$, $\text{p-value}(W_0(1; 2, 3)) = 0.049$, $\text{p-value}(W_1(1; 2, 3)) = 0.049$, the hypothesis of equal APCs is rejected with 0.05 significance level for the three regions when using the Wald-type test, while cannot be rejected using Z -test-statistics for couples of regions.

Acknowledgments

This work was partially supported by Grant MTM 2009-10072.

References

1. Dik JJ, de Gunst MCM. The distribution of general quadratic forms in normal variables. *Statistica Neerl.* 1985; 39:14–26.
2. Horner, MJ.; Ries, LAG.; Krapcho, M.; Neyman, N.; Aminou, R.; Howlader, N.; Altekruse, SF.; Feuer, EJ.; Huang, L.; Mariotto, A.; Miller, BA.; Lewis, DR.; Eisner, MP.; Stinchcomb, DG.; Edwards, BK., editors. *SEER Cancer Statistics Review, 1975-2006*. National Cancer Institute; Bethesda:
3. Martín, N.; Li, Y. A New Class of Minimum Power Divergence Estimators with Applications to Cancer Surveillance. *The Berkeley Electronic Press*; 2009. p. 109Harvard University Biostatistics Working Paper Series
4. Pardo, L. *Statistical Inference Based on Divergence Measures*. Chapman & Hall/CRC; New York: 2006.
5. Pardo L, Martin N. Homogeneity/Heterogeneity Hypotheses for Standardized Mortality Ratios Based on Minimum Power-divergence Estimators. *Biomet J.* 2009; 51:819–836.
6. Li Y, Tiwari RC, Zou Z. An age-stratified model for comparing trends in cancer rates across overlapping regions. *Biomet J.* 2008; 50:608–619.

Fig. 1.
Fitted models to data

Thyroid cancer mortality trends comparison among WC, SW and W during 1969-1983, 1977-1991 and 1990-1999 respectively: Maximum Likelihood Estimators and Minimum Chi-Square Estimators.

Table 1

Region	k	λ	$\beta_{0k,\lambda}$	$\beta_{1k,\lambda}$	$\widehat{APC}_{k,\lambda}$
WC	1	0	-0.3680	-0.0267	-2.639
	1	1	-0.3241	-0.0268	-2.646
SW	2	0	-0.5404	-0.0107	-1.064
	2	1	-0.4943	-0.0106	-1.053
W	3	0	-0.7939	0.0003	0.031
	3	1	-0.7084	-0.0012	-0.117
WC vs. SW: $Z_{2,0} = -1.85$; $Z_{2,1} = -1.92$					
SW vs. W: $Z_{23,0} = -0.85$; $Z_{23,1} = -0.75$					
$W_0(1; 2, 3) = 6.02$; $W_1(1; 2, 3) = 6.01$					